Reliability, validity, and feasibility of the objective structured clinical examination in assessing clinical skills of final year surgical clerkship

Mohammed Y. Al-Naami, FRCSC, FACS.

ABSTRACT

الأهداف: لعرض خبرتنا مع أول امتحان إكلينيكي تركيبي موضوعي – أُوسكي (OSCE) طُبِق على طالبات السنة النهائية في قسم الجراحة.

الطريقة: أجريت ورشة عمل تمهيدية للأوسكي (OSCE)، على طالبات السنة النهائية في قسم الجراحة – كلية الطب – جامعة الملك سعود – الرياض – المملكة العربية السعودية، شملت جميع أعضاء هيئة التدريس بقسم الجراحة وذلك قبل موعد الامتحان بشهر تقريبا، حيث أن معظم الحالات الجراحية مع القائمة الفحصية تم تكوينها خلال ورشة العمل. قمنا بتطبيق امتحان الأُوسكي (OSCE) في 22–23 مايو 2005م، على 64 طالبة منقسمات على يومين متتاليين والذي استمر لمدة 160 دقيقة في كل يوم. كان هناك 24 محطة فعالة في اليوم الأول، 22 محطة فعالة في اليوم الثاني، وفي اليومين تكونت المحطات من خليط من المهارات، منها مهارات الاتصال بالمريض، مهارات الفحص السريري، ومهارات حل المشاكل المتعلقة بالحالات الجراحية المستخدمة.

النتائج: المُعامِلات الدالة على إعتمادية الأُوسكي (OSCE) بمقياس ألفا كرونباخ في اليومين الأول والثاني كانتا 0.68 و0.79 على التوالي. ومعامِلات ثبات الأُوسكي (OSCE) الداخلي بمقياس ثيتا كارمين في اليومين الأول والثاني كانتا 0.79 و0.81 على التوالي. أما مصداقية وشمولية امتحان الأُوسكي (OSCE) فقد اعتبرت جيده بواسطة أعضاء هيئة التدريس والطالبات على حد سواء. أما دقة امتحان الأوسكي (OSCE) والتي قيست بمقارنة نتائجة بنتائج الامتحان النظري متعدد الخيارات باستخدام ترابط بيرسون فكانت بمقدار0.5.

خاتمة: أثبت الامتحان الإكلينيكي التركيبي الموضوعي أُوسكي (OSCE) أنه يمكن الاعتماد عليه، إضافة على المصداقية في تقييم المهارات اللاكلينيكية لطلبة الجراحة بالسنة النهائية. أما بالنسبة لإمكانية إجراء امتحان الأُوسكي (OSCE) فيمكن أن يتطور بالخبرة، التكرار، واستخدام المرضى الوهميين.

Objective: To report our experience with measurements of the objective structured clinical examination (OSCE) conducted for the final surgical clerkship year.

Methods: A pilot study of the OSCE was conducted on 64 students split over 2 consecutive days lasting 160 minutes each day in May 22-23, 2005 at the Department of Surgery, College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia. There were 24 actual stations in day one and 22 in day 2, consisting of combinations of history taking/ communication skills, physical examination skills, and problem solving skills.

Results: The stability of the OSCE measured by Cronbach's alpha on day one was 0.68 and 0.79 on day 2. The internal consistency of the OSCE measured by Carmine's theta on day one was 0.79 and 0.81 on day 2. Credibility and comprehensiveness of the OSCE were considered good by both staff and students. Accuracy of the OSCE measured by Pearson's correlation with the MCQs was 0.5.

Conclusions: The OSCE proved to be reliable, and a valid format for testing the clinical skills of final surgical clerkship year. Feasibility of the OSCE can be further improved with experience, repetitions, and use of standardized patients.

Saudi Med J 2008; Vol. 29 (12): 1802-1807

From the Department of Surgery, College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia.

Received 30th June 2008. Accepted 27th October 2008.

Address correspondence and reprint request to: Dr. Mohammed Y. Al-Naami, General Surgery Division, Department of Surgery #37, King Saud University, PO Box 7805, Riyadh 11472, Kingdom of Saudi Arabia. Tel. +966 (1) 4671585. Fax. +966 (1) 4679493. E-mail: mohammed_alnaami@yahoo.com / alnaami@ksu.edu.sa

Assessment assures clinical competence and drives the learning process.[1-3] The objective structured clinical examination (OSCE) involves several clinical encounters (stations) through which examinees rotate equally and are rated by examiners or standardized (simulated) patients using standardized checklists. Since its introduction by Harden and Gleeson,[4] the OSCE is gaining popularity among medical schools as a strong clinical assessment tool. This has been attributed to strengths in its reliability and validity.[5-7] Feasibility of the OSCE, however, still poses a problem of being resource-intensive.[7] Reliability of an examination refers to its consistency or precision in discriminating students' performance upon repetitions and when examiners are in close agreement in their ratings.[8,9] Reliability is determined by a correlation coefficient (r) that can be measured using the Kuder-Richardson methods,[9] more accurately by the generalizability theory,[10] or currently by using different accurate statistical software programs available commercially. A zero correlation coefficient indicates no reliability and one indicates perfect reliability.[11] Reliability can be further sub-categorized into stability and internal consistency. Stability means that the examination is stable in discriminating students' performance upon repetitions. Ideally, stability correlation coefficient should exceed 0.5.[11] Internal consistency means that scores of an examination in each item (station) would be correlated with scores of all other items (stations). Internal consistency coefficient ideally should exceed 0.8.[11]

Several authors have made different recommendations regarding the minimum acceptable level of reliability. In general, reliability correlation coefficient (r) <0.5 is considered low, (r) ≥0.5-≤0.8 is moderate, and (r) >0.8 is high. An examination with low reliability is usually unacceptable. Moderate reliability is generally acceptable for students' progress (for example continuous assessment, yearly performance, passing a course within the curriculum, and so forth).[6] However, high reliability coefficient (>0.8) is required for critical decision making (for example graduation, board certification, and so forth).[11] Validity is the degree to which an examination is measuring what it is supposed to measure.[8,9,11] Validity of an examination can be expressed by several aspects. Credibility (face validity) reflects the degree to which an examination can measure what it is supposing to measure (for example clinical reality by direct observation of a student interviewing or examining an actual or standardized patient). Comprehensiveness (content validity) means the extent to which an examination can sample from different content domains or topics [for example multiple choice questions (MCQs) are a content valid measure as it can sample from a wide range of topics]. These validity aspects can be measured non-quantitatively by experts' and students' judgment. Accuracy (concurrent validity) implies a statistical association with "best" available or "gold standard" measure or examination (for example correlating students' scores of modified essay questions with scores of MCQs for testing knowledge domain).

Predictive validity refers to the association of an examination with an outcome that can be predicted in the future (for example correlating student's scores in admission test with scores of graduation test). Construct validity means demonstration of expected "theoretical" or "hypothetical" difference in performance using the test in question (for example demonstration of a better performance of senior surgical residents over junior surgical residents in the in-training OSCE).[12] The latter aspects of validity are measured quantitatively by correlation coefficients that estimate the degree to which an examination is truly measuring what it is supposed to measure. The relationship between reliability and validity is very important as a reliable examination must be valid to be acceptable. Feasibility refers to practicality and cost of the examination. Feasibility complements reliability and validity. Therefore, if an examination is reliable and valid, but is too difficult to implement, this examination becomes unacceptable or may be impossible. The purpose of this study is to report our experience with the first OSCE conducted for assessing clinical skills of final year surgical clerkship.

Methods. One month prior to final year surgical clerkship examinations, a half-day workshop on OSCE was conducted at the Department of Surgery, College of Medicine, King Saud University, Riyadh, Kingdom of Saudi Arabia with the approval of our ethics committee, where more than 25 surgical consultants and associate consultants from general surgery and surgical specialties were introduced to the OSCE by a member from the department who is certified in medical education and experienced in OSCEs. The workshop started early morning with an interactive one hour lecture presentation, followed by live demonstration of an OSCE station consisting of pioneered individuals from within the groups who were given their roles as, standardized patient complaining of upper abdominal pain, examiner, and examinee. A prepared checklist sample, the same as the one with the examiner, was projected on the screen containing important items relevant to that station with a scoring method. A mixture of 20 surgical encounters (blueprint) encompassing 10 general surgery scenarios and 10 surgical specialties scenarios (involving neurosurgery, vascular, plastic, anesthesia, urology, cardiac, and thoracic surgery) were proposed and developed along with their checklists by almost all participants during the workshop. At the end

of the workshop, all developed stations were presented by corresponding individuals, discussed, and improved further by discussants to match students' level. These stations were typed and saved confidentially for use and revision if necessary. After this workshop, most of our surgical staff became familiar with OSCE stations development. An OSCE committee was then developed for each surgical course consisting of a consultant course coordinator, a consultant co-coordinator, 2 senior registrars (assistant consultants) and secretary to overlook the final year OSCE and its conduct and to recruit more stations for enriching a future OSCE station bank. The final year OSCE was conducted in May 22-23, 2005 that included 64 medical students split over 2 consecutive days having each student passed through 32 stations each day. The OSCE was conducted in 2 surgical wards with 2 rest stations (5 minutes each) in and out of each ward. Twenty-four rated (actual) stations were used on day one and 22 on day 2, while 8 were unrated (rest) stations on day one and 10 on day 2. Each station lasted for 5 minutes totaling 160 minutes OSCE each day. Movement of candidates from one station to the other followed an audible ring located at the central nursing station in each ward. Almost all stations involved an examiner who developed that corresponding station with its checklist during the preceding workshop. Some unmanned (no examiner) stations involved short answers to x-rays, pictures, instruments, specimens, and so forth. Each examiner marked the station's checklist for each student out of 10 using the global rating scoring method.[6] Results were submitted to the OSCE committee for summation and final checking. Stations' blueprint (desired clinical skills to be examined) involved history taking, and communication skills, physical examination skills, and problem solving skills with an almost equal proportion for each domain.

Stability and internal consistency of the OSCE were measured by Cronbach's alpha and Carmine's theta, using the BMDP® statistical software. Credibility and comprehensiveness of the OSCE were judged by faculty and students. Accuracy (comparing OSCE with MCQ scores) was measured by Pearson's correlation coefficient. Construct and predictive validities were not applicable to this study. Feasibility issues are discussed in the text.

Results. Stations used for final year surgical clerkship OSCE in day one are presented in **Table 1**. Cronbach's alpha (measure of stability) column indicates overall reliability coefficients with corresponding station being removed. This gives an idea on each individual station, if its removal will affect overall reliability of the OSCE. Obviously, all stations are stable and removal of anyone also do not affect overall reliability. Carmine's theta (measure of internal consistency) column indicates correlation coefficients of each individual station's scores to overall OSCE stations scores. If correlation coefficient is high (≥0.8), this indicates a consistent station in discriminating students' performance. Correlation coefficients from 0.5-0.8 are moderate, indicating that the corresponding station can be further improved by minor revisions; however, correlation coefficients <0.5 indicates that corresponding station needs to be removed or totally revised for future use. History taking and communication skills were assessed on 9 stations with overall moderate stability (r) =0.66. Internal consistency correlation coefficients were highest (≥0.8) for gallstones, hematuria, and the inguinal hernia stations indicating that these stations were the most consistent among all other history stations. Physical

Table 1 - Description, stability, and internal consistency of day one OSCE stations.

Station type	Cronbach's alpha (stability)	Carmine's Theta (internal consistency)
History taking and communication skills		
1. Biliary colic	0.6719*	0.8595†
2. Breast cancer	0.6703	0.7617
3. LUTS	0.6385	0.7288
4. Hematuria	0.6701	0.7956
5. Peripheral vascular occlusion	0.7001	0.7760
6. Polytrauma (standardized patient)	0.6289	0.6620
7. Bowel obstruction	0.6764	0.6453
8. Inguinal hernia	0.6287	0.8879
9. Diabetic foot	0.6597	0.6440
Total history	**0.6608**	**0.7512**
Physical examination		
10. Subcutaneous lipoma	0.6709	0.8586
11. Undescended testis (child)	0.6785	0.8234
12. Multinodular goiter	0.6879	0.8292
13. Peripheral vascular occlusion	0.6604	0.7925
14. Incisional hernia	0.6851	0.6830
15. Appendicular mass	0.7048	0.7663
16. Parotid tumor	0.6537	0.6620
17. Thyroglossal duct cyst	0.6666	0.7528
Total physical examination	**0.6760**	**0.7710**
Problem solving skills		
18. Intestinal obstruction x-rays (neonate)	0.6654	0.6834
19. Child with nephrostomy tube	0.6287	0.7303
20. Subdural hematoma CT films	0.6938	0.9091
21. Cleft lip and palate picture	0.6959	0.8649
22. Post-operative bile leak case	0.6684	0.8687
23. Renal stones intravenous urogram	0.6718	0.7263
24. Diverticular disease barium enema	0.6429	0.8749
Total problem solving	**0.6667**	**0.8082**
Overall OSCE Stations	0.68	0.79

*Overall OSCE stability coefficient with this station removed, †Station's coefficient to the sum of all stations' coefficient correlation (item total correlation). OSCE - objective structured clinical examination, LUTS - lower urinary tract syndrome, CT - computed tomography

examination skills were assessed on 8 actual patients (stations). Overall stability was also moderate for physical examination stations. Subcutaneous lipoma, undescended testis, and multinodular goiter stations were the most consistent stations, while the rest had moderate internal consistencies. Problems solving skills were assessed on 7 stations with overall moderate stability. The intestinal obstruction station involved x-rays of a neonatal small bowel atresia case, and a pediatric surgeon examiner interrogating each student using a standardized checklist. The nephrostomy tube station involved a child with a nephrostomy tube for treating congenital obstruction at the pelviureteral junction and a pediatric urologist examiner. The subdural hematoma station involved a CT head showing that pathology and attended by a neurosurgeon examiner; this station scored the highest internal consistency coefficient (>0.9). The cleft lip and palate station involved pictures showing the anomalies and a plastic surgeon examiner. The bile leak and drain station involved an actual patient after complicated gallbladder surgery and examiner. The renal stones and diverticular disease unmanned (no examiner) stations were presented as urogram and barium enema films, with short answer questions and a box for dropping completed answer sheets. Most of the problem solving skills stations had an overall high internal consistency (>0.8).

Table 2 - Description, stability, and internal consistency of day 2 OSCE stations

Station type	Cronbach's alpha (stability)	Carmine's Theta (internal consistency)
History taking and communication skills		
1. Renal colic	0.7646*	0.7464†
2. Bowel obstruction	0.7894	0.8325
3. LUTS	0.7814	0.8028
4. Inguinal hernia	0.7733	0.6979
5. Dysphagia (standardized patient)	0.7895	0.7263
Total history	**0.7796**	**0.7610**
Physical examination		
6. Abdominal examination	0.7968	0.8081
7. Multinodular goiter	0.7845	0.8064
8. Incisional hernia	0.7735	0.7056
9. Breast cancer	0.7810	0.7824
10. Subcutaneous lipoma	0.7729	0.7747
11. Diabetic foot	0.7813	0.5686
12. Polytrauma (standardized patient)	0.7823	0.6734
13. Thyroglossal duct cyst	0.7771	0.6426
Total physical examination	**0.6760**	0.7710
Problem Solving Skills		
14. Congenital diaphragmatic hernia x-rays (neonate)	0.7768	0.6749
15. Colostomy case	0.7657	0.7246
16. Nephrostomy tube case	0.7729	0.8498
17. Brain tumor CT films	0.7730	0.8484
18. Undescended testis (child)	0.7689	0.8145
19. Post-operative bile leak (drain)	0.7719	0.7826
20. Urinary bladder tumor (cystogram)	0.7759	0.7419
21. ERCP film	0.7830	0.6551
22. Burn pictures	0.7795	0.8077
Total problem solving	**0.7742**	0.7666
Overall OSCE Stations	0.79	0.81

*Overall OSCE stability coefficient with this station removed, †Station's coefficient to the sum of all stations' coefficient correlation (item total correlation). OSCE - objective structured clinical examination, LUTS - lower urinary tract syndrome, CT - computed tomography ERCP - endoscopic retrograde cholangiopancreatography

Day 2 OSCE stations are presented in **Table 2**. Overall history taking and communication skills stations have shown higher stability (r=0.78) compared to day one. Overall internal consistency coefficient was moderate, with bowel obstruction and lower urinary tract syndrome history stations scoring the highest correlations (>0.8). Overall physical examination stations' stability and internal consistency measures were moderate, with abdominal examination and multinodular goiter stations scoring the highest internal consistency coefficients (>0.8). Overall stability and internal consistency of physical examination stations were similar to day one. Overall problem solving skills stations scored moderate stability and internal consistency. Nephrostomy tube, CT of brain tumor, undescended testis, and burn picture stations scored high internal consistency correlation coefficients (>0.8). No examiners were involved in the cystography of a bladder tumor and endoscopic retrograde cholangiopancreatography showing biliary stones films. Overall OSCE stations' stability and internal consistency for days one and 2 were comparable without significant statistical difference. Comprehensiveness and credibility of the OSCE were judged as very good by most staff and students. Accuracy (concurrent validity) of overall OSCE scores compared to overall scores of MCQs measured by Pearson's correlation coefficient was 0.5. Construct and predictive validities were not addressed in this study. Feasibility of this OSCE was judged as labor-intensive with lack of standardized patients.

Discussion. Since its introduction in the late seventies, the OSCE was developed into different themes.[13,14] Some authors use actual patients; others use standardized (simulated) patients; and some use other material including investigational results, values, pictures, instruments, and so forth, or a combination of things depending on the required competencies or skills to be tested. We used a combination of stations encompassing history taking and communications skills, physical examination skills, and problem solving skills. Each competency area (domain) constituted one third of the whole OSCE. Initial preparation of OSCE,

which includes staff preparedness and commitment, identifying objective skills to be tested (blueprint), and stations development is of paramount importance.[6,15] In this OSCE, most of these preparatory steps were achieved during the workshop made prior to the actual OSCE. Some of the developed stations were revised and modified as necessary to match clinical findings of patients available during the actual OSCE. Other stations were developed immediately prior to the actual OSCE for cases that have not been prepared previously (for example post operative bile leak and nephrostomy tube cases). All patients were real, except one standardized patient who gave history of polytrauma in day one and underwent polytrauma physical examination in day 2, and one standardized patient who gave history of dysphagia in day 2. Those standardized patients were examiners at the same time. Surgical wards were selected as ideal locations for the conduct of our OSCEs. Empty spaces (for example clinics during weekends, skills laboratory, and so forth) could have been more ideal for OSCE conduct to avoid disturbing other patients and conflict with other activities within the wards.[14]

The standard length of an OSCE station encounter ranges from 5-10 minutes,[6,15] however, acceptable time for the whole OSCE is controversial as some authors have achieved good reliability correlations (≥0.8) with approximately 4 hours OSCE,[16] and others have recommended 6-8 hours for reliable results.[17] In this OSCE, most stations had moderate stability and internal consistency correlations ranging from 0.6-0.8. Although a few standardized patients were used; however, there was no significant difference between actual and simulated patients performance as supported by some authors.[18,19] We faced some difficulties dealing with some actual patients during the OSCE because of their fatigue and non-cooperation. On day 2, some patients actually did not show up, resulting in their replacement with other standby problem solving skills stations. To prevent this from happening in the future, the OSCE committee has decided to use more than one patient to alternate in the same station. Also, more standardized patients have to be recruited and trained for that purpose. Test security has been an issue if the OSCE is repeated for the same cohort group of students on same or subsequent days. This notion, although genuine, has not been substantiated by literature. Some authors have shown no significant statistical difference among cohort groups performance in OSCE across clerkship rotations.[20]

Our results have shown moderate overall reliabilities without a significant difference between day one and 2. We believe that skills that were not acquired during the course cannot be acquired during the same day or overnight. Validity of OSCEs have been reported to be high,[5-7] especially content validity being the most important requirement in establishing a valid examination.[10] The content of our OSCE shows a wide range of clinical and problem solving skills as judged by faculty and students. Credibility (face validity) of the OSCE has been critiqued, as 5-10 minute encounters are fragmented and lacking in-depth testing of knowledge and skills.[3] However, in-depth skill testing requires longer stations' time, which may result in fewer encounters that may affect comprehensiveness of the test. The OSCE seems to be credible at undergraduate medical students' level and in-depth knowledge and skills testing may be more necessary for postgraduate students.[21]

Our OSCE was also considered to be credible and adequate. Concurrent validity of the OSCE requires a "gold standard" clinical test for correlation, which is lacking in many situations.[22] Reported concurrent validities of the OSCE were mostly carried out by correlations with written tests. These correlations are generally low.[23,24] Moderate and high correlations of OSCEs with written tests have been also reported.[25,26] Our OSCE scores have indicated moderate concurrent validity correlation of 0.5 with the written one best answer MCQs test scores. Predictive and construct validities were not addressed in this study as they require further comparisons with past or future performance of the same group of students and with another non-cohort group of students, which make this study a limited one. Most of arguments on the OSCE are related to its feasibility. Many authors have reported that OSCEs are costly and labour-intensive.[7,27] We have experienced some difficulties in our first OSCE. These difficulties were mostly related to lack of proper organization and some patients' non-cooperation.

In conclusion, the OSCE proved to be a reliable and valid format for testing clinical skills of final year surgical clerkship. Feasibility of the OSCE can be further improved with experience, repetitions, and use of standardized patients.

References

1. Frederickson N. The real test bias: influences of testing on teaching and learning. ***Am Psychol*** 1984: 39: 193-202.
2. Norman GR. Current issues in the assessment of clinical competence. In: Hart IR, Harden RM, editors. Further developments in Assessing clinical Competence. Montreal, Que: Can-Heal Publication Inc.; 1987. p. 136-146.
3. Newble DI. 8 years' experience with a structured clinical exam. ***Med Educ*** 1988; 22: 200-204.
4. Harden RM, Gleeson FA. Assessment of clinical competence using an objective structured clinical examination (OSCE). ***Med Educ*** 1979; 13: 41-54.
5. Regehr G, MacRae H, Reznick R, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. ***Acad Med*** 1998; 73: 993-997.

6. Newble DI. Techniques for measuring clinical competence: objective structured clinical examinations. ***Med Educ*** 2004; 38: 199-203.
7. Barman A. Critiques on the objective structured clinical examination. ***Ann Acad Med Singapore*** 2005; 34: 478-482.
8. Neufeld VR. An introduction to measurement properties. In: Neufeld VR, Norman GR, editors. Assessing Clinical Competence. New York (NY): Springer Publishing Company; 1985. p. 39-50.
9. Gronlund NE, Linn RL. Validity, Reliability, and other desired characteristics. In: Gronlund NE, Linn RL, editors. Measurement and Evaluation in Teaching. New York (NY): Macmillan Publishing Company; 1990. p. 47-106.
10. Crossley J, Davies H, Humphris G, Jolly B. Generlisability: a key to unlock professional assessment. ***Med Educ*** 2002; 36: 972-978.
11. Streiner DL, Norman GR. Reliability. In: Streiner DL, Norman GR, editors. Health Measurement Scales. Oxford: Oxford University Press; 1991. p. 79-96.
12. Cohen R, Reznick R, Taylor B, Provan J, Rothman A. Reliability and validity of the objective structured clinical examination in assessing surgical residents. ***Am J Surg*** 1990; 160: 302-305.
13. Hodges B. OSCE Variations on a theme by Harden. ***Med Educ*** 2003; 37: 1134-1140.
14. Nicol M, Freeth D. Assessment of clinical skills: a new approach to an old problem. ***Nurse Educ Today*** 1998; 18; 601-609.
15. Heard JK, Allen RM, Cason GJ, Cantrell M, Tank PW. Practical issues in developing a program for objective assessment of clinical skills. ***Med Teacher*** 1998; 20: 15-21.
16. Stillman PL, Regan MB, Swanson DB, Case S, McCahan J, Feinblatt J, et al. An assessment of the clinical skills of 4th year students at 4 New England Medical Schools. ***Acad Med*** 1990; 65: 320-326.
17. Newble DI, Swanson DB. Psychometric characteristics of the objective structured clinical examination. ***Med Educ*** 1988; 22: 325-334.
18. Wallace J, Rao R, Haslam R. Simulated patients and objective structured clinical examinations: review of their use in medical education. ***Adv Psych Treatment*** 2002; 8: 342-348.
19. Nackman GB, Bermann M, Hammond J. Effective use of human simulators in surgical education. ***J Surg Res*** 2003; 115: 214-218.
20. Niehaus AH, DaRosa DA, Markwell SJ, FolseR. Is test security a concern when OSCE stations are repeated across clerkship rotations? ***Acad Med*** 1996; 71: 287-289.
21. Patil JJ. Objective structured clinical examination. ***Can Med Assoc J*** 1993; 149: 1376-1378.
22. Nayer M. An overview of the objective structured clinical examination. ***Physiother Can*** 1993; 45: 171-178.
23. Skakun EN, Morrison JC, Elloker MG. Using an objective structured clinical examination to assess components of clinical skills in medicine. In: Bender W, Hiemstra RJ, Scherpbier AJJA, Zwierstra RP, editors. Teaching and assessing clinical competence. Netherlands: Stichting TICTAC, Groningen; 1990. 254-260.
24. Merrick HW, Nowacek G, Boyer J, Robertson J. Comparison of the objective structured clinical examination with the performance of third-year medical students in surgery. ***Am J Surg*** 2000; 179: 286-288.
25. Gerrow JD, Murphy HJ, Boyd MA, Scot DA. Concurrent validity of written and OSCE components of the Canadian dental certification examinations. ***J Dent Educ*** 2003; 67: 896-901.
26. Matsell DG, Wolfish NM, Hsu E. Reliability and validity of the objective structured clinical examination in paediatrics. ***Med Educ*** 1991; 25: 293-299.
27. Reznick RK, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, et al. Guidelines for estimating the real cost of an objective structured clinical examination. ***Acad Med*** 1993; 68: 513-517.

Corrections, retractions and "Expressions of Concern"

Excerpts from the Uniform Requirements for Manuscripts Submitted to Biomedical Journals updated November 2003. Available from www.icmje.org

The corrections should appear on a numbered page, be listed in the contents page, include the complete original citation and link to the original article and vice versa if online.

The retraction and expression of concern, appear on a numbered page in a prominent section of the print journal as well as in the online version, be listed in the contents page, and include in its heading the title of the original article. The text of the retraction should explain why the article is being retracted and include a full original citation reference to it.

Editors may ask the author's institution to assure them of the validity of earlier work published in their journals or to retract it. If this is not done editors may choose to publish an announcement expressing concern that the validity of previously published work is uncertain.